WEALTH FROM OCEANS FLAGSHIP
www.csiro.au

SARC 58: Benchmark stock assessments for butterfish, tilefish and northern shrimp

Woods Hole, Massachusetts

27-31 January, 2014

Dr Cathy Dichmont, CSIRO

Prepared for the Center of Independent Experts

Contents

Contents i

Figures ... iii

1 Executive Summary ... 1

2 Background ... 4

3 Description of the Individual Reviewer's Role in the Review Activities ... 5

4 Findings ... 6

4.1 A. Butterfish Terms of Reference ... 6

ToR 1 Characterize the commercial catch including landings, effort and discards by gear type. Describe the magnitude of uncertainty in these sources of data. 6

ToR 2 Characterize the survey data that are being used in the assessment. Describe the magnitude of uncertainty in these sources of data. 7

ToR 3 Characterize oceanographic and habitat data as it pertains to butterfish distribution and availability. If possible, integrate the results into the stock assessment (TOR-5). 8

ToR 4 Evaluate consumptive removals of butterfish by its predators. If possible, integrate results into the stock assessment (TOR-5). 10

ToR 5 Use assessment models to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a comparison with previous assessment results and previous projections. 10

ToR 6 State the existing stock status definitions for "overfished" and "overfishing". Given that the stock status is currently unknown, update or redefine biological reference points (BRPs; point estimates for BMSY, BTHRESHOLD, FMSY and MSY, or their proxies) and provide estimates of their uncertainty. Consider effects of environmental factors on stability of reference points and implications for stock status. 14

ToR 7 Evaluate stock status with respect to a newly proposed model and with respect to "new" BRPs and their estimates (from TOR-6). Evaluate whether the stock is rebuilt. 14

ToR 8 Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs). 15

ToR 9 Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations. 15

4.2 B. Tilefish Terms of Reference ... 16

ToR 1 Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the magnitude of uncertainty in these sources of data. 16

ToR 2 Characterize commercial LPUE as a measure of relative abundance. Consider the utility of recreational data for this purpose. Characterize the uncertainty and any bias in these sources of data. 17

ToR 3 For the depth zone occupied by tilefish, examine the relationship between bottom temperature, tilefish distribution and thermal tolerance. 18

ToR 4 Use assessment models to estimate annual fishing mortality and stock size for the time series, and estimate their uncertainty. Include a historical retrospective to allow a comparison with previous assessment results. 18

ToR 5 State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates for BMSY, BTHRESHOLD, FMSY and MSY or for their proxies) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs. 21

ToR 6 Evaluate stock status with respect to the existing ASPIC model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt. 21

ToR 7 Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs). 22

ToR 8 Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations. 22

4.3 C. Northern Shrimp Terms of Reference 23

ToR 1 Present the Gulf of Maine northern shrimp landings, discards, effort, and fishery-independent data used in the assessment. Characterize the precision and accuracy of the data and justify inclusion or elimination of data sources. 23

ToR 2 Estimate population parameters (fishing mortality, biomass, and abundance) using assessment models. Evaluate model performance and stability through sensitivity analyses and retrospective analysis, including alternative natural mortality (M) scenarios. Include consideration of environmental effects where possible. Discuss the effects of data strengths and weaknesses on model results and performance. 24

ToR 3 Update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, SSBMSY, FMSY, MSY). Evaluate stock status based on BRPs. 27

ToR 4 Characterize uncertainty of model estimates of fishing mortality, biomass and recruitment, and biological reference points. 27

ToR 5 Review the methods used to calculate the annual target catch and characterize uncertainty of target catch estimates. 28

ToR 6 Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made before the next benchmark assessment. 28

ToR 7 Based on the biology of species, and potential scientific advances, comment on the appropriate timing of the next benchmark assessment and intermediate updates. 29

5 Conclusions and Recommendations 30

6 References 31

Appendix 1 Bibliography of materials provided for review 32

6.1 Working Papers 32

6.2 Background papers 32

Appendix 2 Copy of the Statement of Work 35

Appendix 3 Panel membership or other pertinent information from the panel review meeting........ 53

Figures

Figure 1: Northern shrimp CPUE of animals and survey index greater than 22 mm CL over time. 29

1 Executive Summary

On the 27 January and 31 January 2014, the 58th Stock Assessment Review Committee (SARC) panel members reviewed the stock assessments for butterfish (*Peprilus triacanthus*), golden tilefish (*Lopholatilus chamaeleonticeps*) and northern shrimp (*Pandalus borealis*) in Woods Hole, MA. The review panel consisted of the chair Dr. Robert J. Latour (MAFMC SSC and Virginia Institute of Marine Science) and three scientists affiliated with the Center for Independent Experts: Drs Cathy Dichmont (CSIRO, Australia), Stewart Frusher (University of Tasmania, Australia), and Ian Jonsen (Dalhousie University, Canada).

The butterfish, tilefish and shrimp assessment teams and associated working groups (WGs) have achieved considerable advances in characterising the underlying data required for the assessments and describing uncertainty. These groups took great care in fully specifying and testing for different kinds of assessment uncertainty: model structure uncertainty through the use of different models with very distinct assumptions, sensitivity of each model to input and parameter uncertainty, and sensitivity of the models to new data through retrospective tests. This approach should be used as a model for other WGs.

<u>Butterfish</u>

For butterfish, all Terms of References (ToRs) were fully or partially met, and the assessment can be used for management advice. Since the last supported assessment, there have been several advances, including:

- Updating the input data to further include uncertainty,
- Starting the assessment model at a later date to coincide with the start of the observer program, thereby removing the most uncertain data,
- Undertaking extensive work to produce an input value for the NEFSC fall offshore survey catchability parameter and allowing natural mortality to be estimated internal to the model,
- Undertaking novel approaches to investigate the impact of the environment on the available butterfish habitat and therefore butterfish availability to the NEFSC fall offshore survey over time,
- Calculating the relative catchability values of the two survey vessels that have undertaken the NEFSC survey so as to create a single combined index,
- Moving to a modified Age Structured Assessment Program (ASAP) assessment, called ASAP3.

The input data and associated error was well characterised and were accepted without change by the panel. Starting the model later (i.e. using a shorter time series), and therefore only including data where discards were estimated using observer data, is appropriate. This decision addressed previous concerns of the unknown scale and uncertainty of discards prior to the observer program. The calculation of fall survey catchability (q) is novel and addresses a previous issue of what input value to place on the survey q. In the proposed base assessment, the q is an input (as a point prior) and natural mortality (M) is estimated. Although this model structure is unusual, the q value analysis was thoughtfully undertaken, had observational basis and applied assumptions that ensured a conservative approach (since the method assumed the survey cruises undertaken by the *FSV Henry B. Bigelow* were 100% efficient at sampling butterfish). Future work should include using the uncertainty in the input q value. The assessment update since the previously accepted assessment were of great additional value.

The panel accepted the proposed base assessment model, but with a few changes:

- The panel recommended a change to the proposed base model by not including the thermal Habitat Suitability Index (HSI). This body of work was impressive and a great addition to the research question of whether the environmental impacts butterfish availability and therefore the survey index. However, the resultant HSI showed little change over time and therefore its inclusion did not impact the assessment result. For model parsimony reasons, this Index was therefore not included in the final accepted model. Despite this finding, further research into this novel approach is recommended, and this Index should be retained as a sensitivity test to check for potential future changes.

- The base model highlighted the conflicting trends between the fall and spring surveys (as demonstrated by patterns in the survey residuals). For this reason, the final approved model only included the NEFSC fall survey. The fall survey was a better sample of the full butterfish population.

The accepted base model (without the habitat index and the spring survey) yielded low estimated fishing mortality (F) rates, and a relatively healthy stock abundance. There was consensus among the Panel members that this slightly modified base model could be used for management purposes. The results show that overfishing is not occurring and the stock is not overfished.

<u>Golden tilefish</u>

For golden tilefish, all ToRs were met. The review process supported using the assessment results for management. Since the last supported assessment, there have been several advances, including:

- Reconstructing the landings and standardised catch rate series,
- Refining the input age and size structure from commercial size categories,
- Development of age and size structured assessments (SCALE and ASAP).

Considerable effort was put into characterising uncertainty through using different models (ASPIC, SCALE, ASAP), undertaking in-depth sensitivity tests of assumptions and input data of the assessments, and retrospective analyses. A characteristic of this assessment is that it relies on longline catch rate (LPUE) as the index of abundance – there was no independent index of abundance for tilefish. There is some uncertainty in the data that could not be resolved, especially since fishing effort in recent years only covers a small component of the tilefish distribution and is recorded at a reasonably coarse level. More small-scale information would be of great benefit in the future. Furthermore, there is uncertainty in the market size category data that could, over time, be reduced through defining a common definition of market categories. Strengthening the index of abundance by developing an industry-based survey that moves beyond the fishing grounds would be of great additional value. Despite all these issues, the use of this industry-based data was supported because size and age cohorts could be tracked over time.

The final base model was accepted by the Panel, but with a minor change in the method of forward projecting the age structure. The final accepted model results showed that overfishing is not occurring and that the stock has been rebuilt.

<u>Northern shrimp</u>

Despite the rigorous efforts of the assessment team, not all the northern shrimp ToRs were met. The assessment models were not robust to tests undertaken during the review and are not recommended for use in management advice. Since the last assessment, large advances were made especially in the delivery of a new additional assessment model (UME – a size-structured assessment designed for hard-to-age animals such as shrimps). Other advances include:

- Undertaking extensive analysis of constant, time and size-varying M values, and
- Calculating a predation index from diet data.

Extensive and rigorous work was undertaken to characterise sources of uncertainty through the development of three independent assessments (ASPIC, CSA and UME), sensitivity tests and retrospective analysis. The latter two assessments are appropriate model types to apply to short-lived hard-to-age species, such as northern shrimp. However, the new data since the previous benchmark assessments have caused model instability in all the assessment model runs provided.

These assessments were not therefore accepted by the panel. The ASPIC model was unable to capture the variability in recruitment and the panel supported the WGs view that this model is not appropriate for a species such as northern shrimp (given the new data). In the proposed UME base model, the fit to the survey indices and size structure was weak with relatively strong residual patterns. This was the basis for not accepting the UME model. The two-stage CSA model was a good compromise between few model parameters (ASPIC) and reasonably large number of estimated parameters (UME). This model also revealed strong residual patterns to the survey indices. However, when the model likelihood component weightings were changed even to reasonably extreme values, there was little to no change in the total likelihood value, but dramatic changes in the management advice that would be provided. A key requirement of a model is robustness to uncertainty. Of concern here, was that fact that the weighting scheme (an input set of values notoriously difficult to independently obtain) essentially determined management advice. The panel therefore also did not accept the CSA model for use in providing management advice. Since none of the assessments was accepted by the panel, it was recommended that stock status in the interim should be determined using a combination of catches and survey indices.

It should be noted that this result with regard to northern shrimp is not a reflection on the assessment team, but rather a feature of the new data showing a large recruitment pulse that is not consistently followed through into the subsequent catch and survey data, thereby providing incongruous data to the assessment. Future work should concentrate on adding potential covariates to the model e.g. bottom temperature.

2 Background

The 58th Stock Assessment Review Committee (SARC) met in Woods Hole, MA between 27 January and 31 January 2014 to review the stock assessments for butterfish (*Peprilus triacanthus*), golden tilefish (*Lopholatilus chamaeleonticeps*) and northern shrimp (*Pandalus borealis*). The review panel consisted of the chair Dr. Robert J. Latour (MAFMC SSC and Virginia Institute of Marine Science) and three scientists affiliated with the Center for Independent Experts: Drs Cathy Dichmont (CSIRO, Australia), Stewart Frusher (University of Tasmania, Australia), and Ian Jonsen (Dalhousie University, Canada).

Support was provided by the Drs James Weinberg (NEFSC SAW Chair) and Paul Rago (NEFSC), and the stock assessment modellers, and associated staff including the relevant WG chairs in the case of butterfish and tilefish. Review and supporting documentation was provided up to two weeks prior to the SARC 58 review. Review documentation was prepared by the NEFSC Coastal/Pelagic Working Group (WG) for butterfish, the Southern Demersal WG for tilefish and by ASMFC Northern Shrimp Technical Committee members with NMFS and University of Maine staff. Presentations were made by Drs. Charles Adams, John Manderson and Timothy Miller (NEFSC) for butterfish; Dr. Paul Nitschke for tilefish; and Ms. Kelly Whitmore and Drs. Anne Richards and Katie Drew for shrimp.

3 Description of the Individual Reviewer's Role in the Review Activities

Within two weeks of the SARC 58 review, assessment review and background documents were made available on an ftp server on the NEFSC website. Documents were easily accessed. The review panel met Drs Weinberg and Rago on the morning prior to the review meeting to provide background regarding the meeting agenda, the panel reporting requirements, logistics and any questions that the panel had about these topics. Drs Weinberg and Latour opened the meeting. The Monday was devoted to butterfish (presentations, questions and discussion), with tasks assigned to the team by the end of the day. The Tuesday was devoted to tilefish and northern shrimp using a similar format to the previous day. Follow-up discussions, particularly over the results of the test runs, were undertaken on the Wednesday and some of Thursday. Thursday was devoted to providing conclusions and editing the three stocks' Assessment Summary Reports.

It should be noted that throughout, the respective teams were enthusiastic and undertook the tasks requested of them with great professionalism and dedication. Some of these tasks were non-trivial. There was time provided for a constructive exchange of ideas and dialogue between the panel members, the WG scientists, other NEFSC and MAFMC staff and industry representatives.

4 Findings

The details of the findings against each Terms of Reference and Species group are described below.

NOTE: BOLD COMMENTS ARE STATEMENTS OF SUPPORT OR RECOMMENDATIONS.

4.1 A. Butterfish Terms of Reference

Background

Butterfish is a short-lived species that typically has a maximum life span of two to three years. There is a single stock from Cape Hatteras to the Gulf of Maine. They are primarily pelagic and in response to seasonal changes in water temperature, migrate inshore and to the north in summer, and offshore and to the south in winter. Their optimal temperature range is between about 4° and 22°C. They spawn in late spring and summer with a peak in June and July. They are also characterised by forming loose schools with a diel vertical migration. Butterfish start maturing when they are age 1 and are fully recruited at age 2. Various elasmobranchs, marine mammals, teleosts and seabirds eat them. They have mainly been caught as a bycatch of the squid fishery.

ToR 1 Characterize the commercial catch including landings, effort and discards by gear type. Describe the magnitude of uncertainty in these sources of data.

<u>This ToR was met</u>

The landings data from this fishery are derived from several sources:

- Prior to 1965 – from Lyles (1967) and compiled in Murawski *et al*. (1978),
- 1965-1989 NEFSC commercial fisheries state canvas data,
- 1990-2012 NEFSC commercial fisheries detail species data tables,
- 1963-1982 and 1983-1986 foreign fleet landings from Waring and Anderson (1983) and NEFSC (1990) respectively. These catches are likely to be underestimates since Spain and Italy did not report their butterfish bycatch from their squid fisheries during the 1972 to 1976 period.

For a large part of the time series (e.g. 58% during 1989-2001 and 67% during 2002-2012), butterfish was a bycatch of the longfin squid fishery. In January 2013, a directed fishery was started. Discards were an important component of the catch and were estimated from the Northeast Regional Office Vessel Tracking and Reporting System and NEFSC Observer Database System (1989 onwards). Since the latter of these discard data sources are seen as more reliable to estimate discards, and the earlier foreign landings series is considered not to be reliable, the catch time series for subsequent ToR analyses starts in 1989. The reasons for using 1989 onward data are:

- No CVs for the foreign catch which occurred mostly prior to 1989, and
- More importantly, an observer program for bycatch started in 1989.

This decision to only use data since the observer program is supported.

Since the last SARC review, a large body of work has focussed on constructing a catch time series and, using the available data, the analyses seem appropriate for subsequent use. The CV's of the post-1989 discard and catch data are still large, but unlikely to be reduced.

Butterfish samples and lengths have been collected for the period 1989-2012 from dockside landings as part of the NEFSC commercial sampling program. Discard size data are sourced from the observer data and have moderate to high CV's.

The observer program is an essential component of estimating the butterfish catch, age and size structure and requires on-going support at present or (if possible) higher (given the high CV's) levels. The observer coverage is also skewed toward the NE area.

Commercial catch-at-age show that the dominant age of the butterfish in the landings are 1 and 2 year olds, whereas the discards are 0 and 1-year old fish, reflecting the short-lived nature of the species.

The recreational catch was insignificant and is **correctly ignored in the analyses**.

Given that the bulk of the catch was bycatch from other fisheries, effort is not well defined and not used in subsequent analysis. **Since the effort is a function of targeting another species, the lack of use in subsequent analysis is supported.**

Catch curve analysis show total fishing mortality (Z) values from about 0.5 to 2.5.

ToR 2 Characterize the survey data that are being used in the assessment. Describe the magnitude of uncertainty in these sources of data.

<u>This ToR was met</u>

Several survey data sources are available for butterfish as listed in the report, however the surveys are designed for groundfish (whereas butterfish are pelagic). In the previous assessment, survey indices from the NEFSC bottom trawl surveys for the winter, spring and fall surveys over the periods 1992-2007, 1973-2008 and 1975-2008 respectively were used. In the spring surveys, only certain inshore and offshore grids were used. The present assessment chose to use only the spring and fall surveys listed in Tables A2.18 and A2.19 respectively. There are several State based surveys which are not included in the assessment model.

Correlation coefficients between the different surveys, including state surveys, show several surveys produce the same relative indices, however many of these are not significantly correlated. Although it is clear that the spatial coverage of the state surveys is limited and may therefore not each reflect overall population abundance, **there seems to be a great deal of potential in applying statistical models (e.g. time-series methods, additive models, hierarchical models, spatial models) to develop a combined index with the NEFSC surveys.**

The NEFSC spring survey indices are affected by a change of survey vessel from the *Albatross IV* (1989-2007) to the *Henry B. Bigelow* (2008-2012). The larger Bigelow is unable to sample the inshore zone compared to the *Albatross*, thereby either requiring calibration of the two data sets to create a single index or treatment of these as separate indices with different catchabilities in the assessment. Unlike some of the other assessments, this assessment calibrates the surveys into a single series. Again, it is unclear why this choice was made (especially since this is not undertaken for other species) or whether it is appropriate. The estimated precision of the NEFSC survey abundance is poorest for the spring series.

For this SARC 58 review, the survey indices have been divided into inshore and offshore indices.

Age data from the NEFSC surveys show that the survey catches age groups 1-3 in spring and 0-3 in the fall.

NEAMAP spring and fall survey data are also presented. The NEAMAP spring abundance indices were higher than the comparable NEFSC spring inshore abundance indices. There was a significant positive correlation between the fall NEFSC offshore and NEAMAP surveys (0.54)

During the review, the assessment team was asked to produce correlation coefficients between the spring and fall survey index, and age cohorts within a survey series.

The correlations between the summer and fall surveys show that the NEAMAP spring and fall surveys are positively correlated (which is a desirable feature). However, the NEFSC spring and fall offshore and inshore surveys are negatively correlated (although not significantly), which means they provide opposing signals of relative abundance over time – an important feature affecting the assessment as discussed in the relevant ToR below.

Correlations that follow age classes over time for the different surveys, show: NEFSC spring ages 1 and 2 are slightly positively correlated; and the NEFSC fall survey age structure tends to be slightly negatively correlated with the subsequent year.

Thus, each survey has some important positive and negative features affecting the assessment:

- The fall survey animals are more evenly distributed within the survey area, but this survey has correlations among cohorts through time that were very weak,
- The spring survey animals are confined to the outer shelf on the edge of the survey sampling frame, so there is some doubt whether the whole population is sampled (which is hinted at by the low numbers of animals sampled). However, it demonstrates the expected positive correlations in cohorts through time.

The magnitude of the uncertainty, especially in terms of CVs and correlations, are well described.

ToR 3 Characterize oceanographic and habitat data as it pertains to butterfish distribution and availability. If possible, integrate the results into the stock assessment (TOR-5).

<u>This ToR was met</u>

This section describes the method used to develop a time varying estimate of availability of butterfish to fishery surveys for the assessment model. Availability is appropriately defined. Availability is one of the two components of catchability. The other being detectability by the trawl gear.

Three reasons for calculating availability are 1) that the assessment relies entirely on survey data, 2) the catch data are from a bycatch fishery and catches have been very low in the recent years – so survey catchability is inestimable in the assessment – and 3) that changes to temperature in recent years may be affecting the stock requiring time-varying catchabilities.

The analysis undertakes a 5-step process and builds a habitat-based index of availability.

It should be noted that this reviewer is not an oceanographer and therefore has no in-depth knowledge on the oceanographic component of the modelling framework. Despite this, the report for this crucial body of work is not well described **and work should go into trying to explain the methods in more precise and clear ways**.

The first step is to develop a debiased bottom temperature hindcast. This process uses a combination of data sources to interpolate bottom temperature over the time and space, relevant to the survey indices, e.g. daily ROMS Bottom temperature and monthly MOCHA bottom temperature climatology data. Uncertainties are also calculated using three temperature states to bound the Index.

The second step developed a thermal niche response to temperature. This step is itself a multi-step process of data mining for an underlying temperature-survey catch response with a statistical model that does not assume an underlying functional shape (i.e. a Generalised Additive Model – GAM). The second is using the closest known temperature response function implied by the GAM, and fitting this function to the data. The GAM pointed to a parametric thermal niche model that was a unimodal extension of Boltzmann-Arrhenius function (Johnson and Lewin, 1946), which has a basis in enzyme kinetics and metabolism. Parameter estimates of this model were obtained from the calibration data by minimising the negative binomial likelihood of the equation using the standardised catch densities as the dependent and bottom temperatures as the independent variable.

The use of this equation, starting values and negative binomial likelihood is justified by the authors through a GAM based on the data that shows an asymmetrical temperature response and the appropriate distribution assumption. For model convergence (in both cases), eight tows with catches

>30,000 fish were removed. **This exclusion of the data is not justified and seems inappropriate given this is a schooling species, and these schools are an important contribution to the survey index.** The base GAM uses (log transformed) swept area as an offset, with a bottom water temperature spline and survey, season and year as factors to estimate butterfish catch. The degree of smoothing was determined internally to R using the *gcv* argument in the GAM function. Different constructs of the model were tested in terms of distribution assumptions and which factors to include. These tests showed that temperature explained 32% of the total deviance, and followed by survey and year effects. The model with the lowest Akaike Information Criterion (AIC) included the survey temperature response, and independent factors survey and year. The calibration data was pooled after the analyses and indicated that dependencies of catch to temperature on survey, year and season were reasonably small.

Both the GAM and ML parametric model does not fit the data well and seems a weak component of the analysis. Although most of the variance is in the centre of the temperature range, no sensitivity to removing more or less of large survey catches are presented to show that the model is reasonably robust. Temperature only explains 32% of the total deviance in the GAM yet it is argued that temperature is the primary response for butterfish.

Both the scaled and unscaled fits to the data by the GAM are provided, but only the scaled graphs for the parametric model. **Both scaled and unscaled version of the model fits should be shown for clarity.**

The third step evaluates the thermal niche model and the thermal niche model coupled to debiased temperature hindcast using data not used in the calibration. A GAM is used to estimate probability of occupancy, with the results showing sensitivity to the threshold for the number of individuals used as the absence threshold. The Habitat Suitability Index (HSI) predicted using bottom temperatures measured in situ fit the data very well, except that false negatives tend to occur in the south and in the inshore. Evaluation of the coupled thermal niche model shows that this model performs extremely well. Despite being better than the uncoupled model, still produces false negatives in the south and inshore.

This step demonstrates the value of this work, as it tends to predict habitat suitability well. **Given the inshore and offshore migration of butterfish, further work should concentrate on the false negatives, particularly within the inshore. This may mean using other surveys, including those from the State.**

The fourth step calculates the habitat based availability index with uncertainties. Integration of the uncertainties from both the niche model and bottom temperature hindcast process is undertaken to then calculate availability using the HSI from each realization and bottom temperature states. This is then used to calculate the median and confidence intervals for the availability index for mean, cold and warm bottom temperature states. **This uncertainty integration is appropriate.**

Step five is an empirical estimate of NEFSC/NEAMAP fall surveys availability ratio over the period used within the assessment. The surveys do not overlap in space – the one being offshore of the other – but do overlap in time.

The NEAMAP survey samples during the daytime whereas the NEFSC samples day and night, so a diel adjustment was made. An underlying assumption was that the daytime NEAMAP survey efficiency is 1 (i.e. all butterfish are sampled). **Although this assumption is unlikely, this is a very conservative assumption that will flow to a more conservative assessment. Given the uncertainty, this assumption is supported.** Flowing from this calculation, the remaining ratios can be calculated empirically. When these ratios are compared to the ratios implied by the habitat index model, they fall within the confidence bounds of each other. A calibration of the transition from the Albatross to the Bigelow is also calculated.

The final Habitat Suitability Index (HSI) is surprisingly time invariant.

It should be noted that the number of variables/parameters implied by the external (to the model) analysis is reasonably large. This is discussed at ToR 5.

In summary, this is an innovative, extensive and rigorous approach to understand butterfish habitat preferences and is a crucial component of the butterfish assessment. It should be continued into the future (despite the HSI not being accepted as the base model for parsimony reasons). It does require further refinement especially with regard to the thermal niche butterfish response to temperature, and the basic assumption that butterfish distribution is largely structured by temperature (when it only explained about 30% of the variance in the GAM). One option is to include other possible variables, especially those relevant to inshore distribution such as chlorophyll.

Given the contrast between the spring and fall survey, it would be of value to calculate the HSI for the spring survey as well.

ToR 4 Evaluate consumptive removals of butterfish by its predators. If possible, integrate results into the stock assessment (TOR-5).

This ToR was partially met

Fish diet data from the NEFSC bottom trawl surveys were investigated, in terms of diet composition of butterfish, total consumption, per capita consumption, amount of butterfish removed by fish predators, and total butterfish consumption by all fish predators. Sample sizes were quite low. Six predators of butterfish were chosen as these consistently ate butterfish. This consumption was consistent and showed no real trend, thus providing support for a constant mortality assumption within the assessment. The reasonably large temporal variability in what eats butterfish was related to the interplay between population size of predators and butterfish.

During the workshop, further analysis of the butterfish data was undertaken. Butterfish consumption was not related to predator or butterfish abundance. This is despite quite large variations in the predator and butterfish abundance during that time.

Given the low samples of the diet data, there would be value in investigating the role of butterfish in the ecosystem using already existing ecosystem models and multi-species modelling. The fall survey has a long-term downward trend, yet the fishing pressure is very low pointing to some potential ecosystem effects.

It is also recommended that these ecosystem and multi-species models be used to look further into the constant M assumption and whether there is some indication of the appropriate scale of M.

ToR 5 Use assessment models to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a comparison with previous assessment results and previous projections.

This ToR was met

The previous two butterfish assessments (NEFSC 2010, 2004) used a delay difference model (an approximate age structured model) and KLAMZ, which tracks recruits and older animals, assumed knife edged selectivity with recruits being fully selected. The last assessment had difficulty estimating the scale of the population without additional information on catchability of butterfish for at least one of the survey indices. They used four different survey indices, which covers the years (at earliest) 1973 and (latest) 2008.

The current assessment used a modified/augmented version of ASAP, which is a statistical catch-at-age model developed by Legault and Restrepo (1999). ASAP forward projects numbers-at-age assuming that fishing mortality can be separated into year and age components using observed catches, catch-at-age and indices of abundance. Observations of proportions at age are modelled assuming a multinomial distribution, while all other model components are assumed to have lognormal error distributions.

In this assessment, different survey data sets are used (see ToR 2) and only data from 1989-2012 were used.

An analysis was undertaken to bridge between the previous and current model. The standard ASAP model is used here. The KLAMZ model was updated to include data to 2012. There are only slight changes in the population size. Using ASAP with the updated set also provided reasonably similar biomass values to KLAMZ, except at the early and most recent periods of the assessments. These results are used as an argument to convert to ASAP as the model of choice.

The reasons for moving to ASAP3 were numerous, including being able to use age data. The move to ASAP3 as the base model is supported.

The ASAP3 augmentation investigated combinations of:

- External or internal (to the model) length-based calibration of survey data in 2009-2012 using the estimates of Miller (2013) as a penalty,
- Estimating or input natural mortality (M), since catchability is an input parameter,
- Including/excluding habitat-based availability

Apart from the different catchability parameters of the two survey vessels, they also have different size effects on their relative catch efficiency. To incorporate uncertainty in the size-based estimates of relative catch efficiency in the assessment model, a penalty is added to the likelihood. This allowed the assumption that the CVs of the indices and effective sample size of the proportions-at-age were the same as if the AIV were being used throughout the series. ***This is correctly implemented.***

Annual or age-specific effects of covariates in M are also estimated or specified. If the catchability parameterisation that constrains catchability is used, then M can be estimated internally.

The estimates of relative survey catchability by length shown in Figure A5.2 have large errors for the start and end of the series. The proposed base model (ASAP3 + M estimated + Habitat-based availability + internal length-based calibration) changes the relative slope relative to that which was calculated. Although there is some departure from normality of the aggregate catch fits to the model, as stated in the text, there is no consistent trend over years. **However, unlike what is stated in the text, there is departure from normality and strong annual trends in the residuals of the model fit to the spring offshore survey (with more of the negative residuals in the early parts of the series – 10 out of the first 12 years – and positive residuals in the latter part of the series – 10 out of the last 11 years).**

This should be corrected in the text as these residual patterns point to a crucial issue for the assessment and form the basis for tests undertaken during the review.

There is also a degree of temporal trend in the NESFC fall offshore survey residuals especially in the early period of the index. **This contradicts statements made in the text and should be corrected**.

Since the two survey series are negatively correlated, the model is unable to fit either of these appropriately and therefore the model fit is an averaging between the two surveys. Leaving this contradiction in the model (when this is the key data source and the model is used for management advice) is not supported.

There is quite strong deviation from normality for the NEFSC fall inshore survey with more negative residuals i.e. the distribution is positively skewed.

Sensitivity tests showed little difference between including the HSI or not. Given the amount of parameters implicit in this index, the rule of parsimony would point to not using the HSI. Its inclusion in the final base model was not supported.

Since it is important to track environmental impacts such as temperature given how much the Gulf of Maine has recently changed, this "+H" model should be included as a sensitivity test.

When compared to the proposed base model, ASAP3+C model run provided:

- **The best total objective function fit,**
- **Similar model fits to the indices, catch and relative efficiency,**
- **Better fit to the catch-at-age composition, and**
- **Only slightly worse fits to the survey age composition.**

The "+C" setting is supported within the approved base case.

A further set of sensitivities were undertaken in which the base model was compared to:

- all spring survey data excluded, or
- assumed full selectivity for all surveys except age 0 for spring surveys.

Unsurprisingly, given the apparent contraction between the NEFSC spring and fall survey indices, the largest impact on model output is the removal of the spring survey data.

A third evaluation of different input natural mortality values had the expected impact. Interestingly, some of the low M values resulted in potentially implausibly high F values for some years. Likelihood profiles of M show larger M values are more appropriate.

A fourth set of sensitivity tests explored assumptions about the catchability of the NEFSC fall survey by fitting models with catchability values ranging from 0.1 to 0.3, while still keeping the habitat-based measures of availability. As expected this test has large impact on the output, with the profile likelihood on catchability favouring lower values.

The final test compared the base model to that using a penalised deviation of the catchability of the NEFSC fall offshore survey. This change had little impact on the results, however including the penalty was important because it included uncertainty in the efficiency estimate in the model.

Tests from generated parametric bootstrapped data showed the base model to be unbiased. See below for additional analysis during the review.

Retrospective analyses of the base model showed a trend in the terminal year estimates of SSB, R and F prior to the inclusion of 2012 data with this impact greatest on R. The scale of these differences was reasonably small.

The assessment team undertook several sensitivity tests during the review. These were undertaken with enthusiasm including adding further runs for our information and ultimate benefit.

These new tests were, running the base model with:

Alt. 1) Spring only indices,
Alt. 2) Fall only indices,
Alt. 3) Fall only indices minus the HSI.

Since the assessment is only estimable with q as an input, and only the q for the fall surveys were calculated, the first test required some assumption for spring q. Since time did not allow this comprehensive calculation, as a short cut the spring survey q (implied by the proposed base model) was used as input in the spring only test. This test was needed to contrast the effect of including all combinations of the surveys, especially excluding the fall survey. The assessment team was also asked to provide likelihood and model fit statistics and to calculate stock status indices.

It is recommended that the basic model sensitivity tests and base model should as a norm show results of the stock status information.

The spring only model runs show much reduced residuals and far fewer patterns (although still not good) in the survey residuals compared to the proposed base model.

Similarly, the survey residual patterns are improved relative to the proposed base model for the NEFSC fall only model run.

The Panel did not accept Alt. 1 because fitting the model required using the proposed base model estimated M as an input. Reservations about the base model precluded this approach.

The Panel felt that Alts 2 and 3 were both viable base configurations, and accepted Alt 3 since (as expected) the model estimates were insensitive to inclusion of the habitat index and the principle of parsimony would therefore favour not including it.

The assessment team was also asked to run prospective tests. These show that the new base model shows little to no pattern in prospective tests.

In summary:

- **The philosophy of inputting catchability due to the low catch is clearly unusual. The Working Group and assessment team has undertaken extensive work and modifications since the previous accepted assessment to mount a convincing case in support of this. In reality, most models either input M or q. Which value of M to use is contentious and often based on very little information. In reality, for this assessment, the basis for the q value is much more rigorous than that for M. Furthermore, the fact that the q value is conservative (and therefore the abundance is conservative) is also important. The basic design behind this model of q being an input and M being estimated is supported.**
- **The proposed base model is not supported, based on:**
    - **Negative correlations between the spring and fall surveys,**
    - **The very strong residual patterns in the proposed model fit to the survey data,**
    - **The lack of robustness of the management proxies and stock status measures due to using different combinations of the fall only, spring only or both surveys in the model (see ToR 6 below for further articulation).**
- **There are strengths and weaknesses of the spring only and fall only model runs:**
    - **Spring only model:**
        - **The survey distribution may not sample the whole population or sample a constant proportion of the population (an essential assumption of its use in the assessment, especially as the only index of abundance)**
        - **Yet the spring only runs had better model fit statistics**
        - **There is little basis for the assumed survey q, since it was a result of estimates from the proposed base model that has not been accepted.**
    - **Fall only model:**
        - **The survey may not sample the inshore well,**
        - **Despite the above statement, it samples most of the population distribution,**
        - **There is no strong residual pattern over time, so this achieves better model fit statistics although still some clustering of positive and negative residuals,**
        - **The basis of the q is accepted by the review panel.**
    - **The model ASAP3+M-H+C-S (fall only) is recommended:**
        - **M should be estimated,**
        - **Given the ability to rigorously develop a catchability parameter, the survey distribution and the better model fit statistics, the base model should be the fit to the fall survey only data rather than averaging between two sets of conflicting survey data,**
        - **There is little sensitivity to adding the +H index, so parsimony dictates that this should not be part of the part of the base model. It is recommended as a sensitivity test as it is a valuable addition to the modelling exercise.**
        - **The +C penalty is appropriate,**
        - **The spring survey should not be included (-S).**
    - **However, in the future:**
        - **The uncertainty in the input q should be included in the assessment, and**

- **The spring q and HSI should be calculated (if possible).**

ToR 6 State the existing stock status definitions for "overfished" and "overfishing". Given that the stock status is currently unknown, update or redefine biological reference points (BRPs; point estimates for BMSY, BTHRESHOLD, FMSY and MSY, or their proxies) and provide estimates of their uncertainty. Consider effects of environmental factors on stability of reference points and implications for stock status.

<u>This ToR was met</u>

From the previous assessment in 2009, it was determined that using equilibrium values was not appropriate given the continual decline of biomass over the entire time series of the model. This means there was no existing peer reviewed and accepted biological reference points for this panel to review.

An alternative Fmsy and Bmsy proxy was therefore proposed with the Fmsy proxy being 2/3M, the M being based on the proposed base model estimated M.

There is little investigation of alternative proxies.

However, this proxy is conservative and a good basis for management. The SARC 58 panel accepted this method of calculating the F proxy, but it was updated using the accepted base model.

The SSBmsy proxy is based on a 50-year projection of the above estimated Fmsy proxy. This is an appropriate method to calculate SSBmsy as it is internally consistent. It was updated with the new Fmsy proxy values.

The proposed BPRs are 0.5Bmsy proxy and Fmsy for overfished and overfishing proxies. This recommendation by the working group and assessment team is supported using the values from the SARC 58 accepted base model.

Given the short-lived nature of the species, environmental factors are likely to affect recruitment estimates and subsequent SSB calculations. The proposed base model explicitly includes the effect of bottom temperature changes on the survey indices through an availability model. These should be captured in the uncertainty around the recruitment estimates and incorporated in the forward projection methodology. This is appropriate, but see comments below for the new BRPs.

Uncertainty is adequately promulgated through the model to the BRPs, although some work on including uncertainty in q should be considered.

ToR 7 Evaluate stock status with respect to a newly proposed model and with respect to "new" BRPs and their estimates (from TOR-6). Evaluate whether the stock is rebuilt.

<u>This ToR was met</u>

The SSBmsy proxy is reasonably sensitive to whether both surveys, fall only or spring only, are used. For this ToR, the updated base model was used to recalculate the BRPs. These show that the SSBmsy proxy is sensitive to the choice of survey data.

Under the proposed and new base models, fishing mortality is low and spawning biomass is healthy in the recent years. The fishery is not overfished or subject to overfishing.

ToR 8 Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

a. Provide numerical annual projections (2 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment is considered (e.g., terminal year abundance, variability in recruitment). Comment on which projections seem most realistic.

b. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

<u>This ToR was met</u>

Stochastic projections were made, assuming recent patterns of fishery selectivity, discarding, maturity-at-age. Future recruitment at age 1 was randomly generated from the pdf of the updated recruitment series of 1989-2012. This method of promulgating the uncertainty in the model to the forward projections is supported.

The proposed method of using the current F of 0.2 for 2013 and Fmsy proxy of 0.85 thereafter was not supported. This is because management decisions for 2014 have already been made and are lower ABCs than the Fmsy proxy value implies.

The Panel accepted a new forward projection method of assuming that existing management decisions would be met i.e. current F for 2013, the ABC for 2014 and Fmsy proxy thereafter.

The SARC 58 accepted base model projected probability of butterfish spawning stock biomass falling below the overfished reference point of 1/2Bmsy, based on 100,000 bootstrap realisations in each year, is <1%.

ToR 9 Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

<u>This ToR was met</u>

The WG reported on progress made against previous research recommendations. Impressive research advancements had been made, most notably in developing q values and a thermal habitat index. Further advances in applying new assessment techniques and new BRPs, and increasing observer coverage to improve discard estimation were made.

The WG also provided four new research recommendations:

- Field experiments on survey catchability and efficiency,
- Information on potential butterfish spawning south of Cape Hatteras, NC,
- Continued improvements on the ASAP3 model, especially with regard to environmental covariates, and
- To not conduct additional assessments unless a fishery has developed that could influence stock size.

The recommendations are supported. However, further research is recommended on:

- **Determining a q and HSI for the NEFSC spring offshore survey,**
- **Expanding the ecosystem research to include the various available Gulf of Maine ecosystem and multispecies models,**

- **Undertaking statistical analysis to fully integrate all available survey data to produce a modelled space-time index of abundance (or something similar),**
- **Re-analyse the HSI GAM to include large schools. Include an investigation of other correlates, beyond temperature, to re-analyse the basis of the temperature-response relationship, and**
- **Undertaking further research on the HSI that concentrates on the reasons for the false negatives, particularly those in the inshore region.**

4.2 B. Tilefish Terms of Reference

Golden tilefish, *Lophatalitus chamaeleonticeps*, are distributed along the upper slope of the continental shelf in depths of 80-440 m from Nova Scotia to South America. They are commonly found in waters ranging from 9 to 14°C. They construct burrows and are therefore not well selected by trawl gear, which means there is no survey-based index of abundance for them. This had an important influence on the assessment. They are long-lived, growing to 39 and 46 years for males and females respectively.

A FMP was implemented in November 2001 with a constant quota of 905 mt. This model was based on a surplus production model, called ASPIC. SARC reviews in 2005 and 2009 concluded that stock rebuilding was on schedule.

ToR 1 Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the magnitude of uncertainty in these sources of data.

<u>This ToR was met</u>

Landings data are available from 1915 to 2012 and range from about 125 mt to over 3900 mt. Catches prior to the mid-1960s were treated as a bycatch from the trawl fishery. The directed longline fishery for tilefish developed in the 1970s, largely in the NJ area. Presently the bulk of the landings come from two statistical areas.

An annual quota of 905 mt was implemented in 2001.

Landings data are characterised as being collected by a series of different data collection systems or methods. Landings from 1915-1972 are from reconstructions in Freeman and Turner (1977), from 1973-1989 are from general canvas data, 1990-1993 from the Weighout system, 1994-2003 from the dealer reported data, and 2004-2012 from dealer electronic reporting. In the early part of the series, there are three years of data missing (1934, 1936, 1941). Until the end of 1972, the location of the landings by statistical area was mainly unknown, whereas thereafter it was divided into area. This has changed over time, especially with regard to area 526. Occasionally, other areas pick up larger proportions of the catch, but mostly they remain below 10%. By State, the results are more consistent with two States dominating the catch over the past few decades. The directed longline is the major gear used since the switch from trawl after 1972.

Discards were estimated following the SBRM approach, which is supported. It also seems reasonable to conclude that the discards from other fisheries since 1989 (based on observed trips) were low, similarly for the directed longline fishery. It is supported that the recreational catch is low.

A lot of thought has gone into reconstructing the best catch series. Based on the comparisons of how landings are arrived at using the different data sources, including investigating the VTR and IVR top 5 vessels, the decisions made are supported.

Landings by market category are available after 1990, length sampling from 1995, age data from 2007. **Uncertainty in the data was well described.**

ToR 2 Characterize commercial LPUE as a measure of relative abundance. Consider the utility of **recreational** data for this purpose. Characterize the uncertainty and any bias in these sources of data.

This ToR was met

There was no independent index of abundance so the assessment relied on landings and effort data to produce a relative index of abundance. **The analysis is correctly limited to the longline fishery for the assessment.**

There are three different series of effort data: tilefish longline effort from 1973-1982 (Turner 1986), NEFC Weighout system (1979-1993), and the VTR system (1995-2013). There was some variability in the use of longline gear in the past, but presently these are consistent especially with respect to number of hooks and number of miles for the top 5.

The catch rate data are limited to trips that had more than 74% tilefish in the total catch (most of the trips are greater than 95% tilefish). The effort unit is calculated as days absent minus number of trips (while also accounting for travel time). Unlike many other longline fisheries, this fishery targets mainly tilefish.

The current unit of effort in use in this fishery is not the best available for longline fisheries. Although this cannot be retrofitted, it is recommended that moving to "hook" or other more traditional longline effort units be investigated.

A simple GLM is undertaken to standardise the effort over time with vessel and year as factors. No within year, or year and season interactions were considered. **Although interactions within the GLM analysis were not tested, the available data presently limit what can be achieved.**

Potential issues with the standardised effort data are that the number of vessels that target tilefish had declined over time to quite low numbers, where five vessels account for more than 70% of the catch. The length of a trip had been generally increasing until the mid 1990s. In recent years, the fishery is slowing down the rate of catches to spread the catch over the year and not flood the market. This has kept prices high in recent years. **It is therefore unlikely that many of the above changes could be standardised by the present data and standardisation. Future work should concentrate on collecting important information needed for standardisation such as gear and vessel characteristics.**

Since currently four vessels catch the bulk of the catch and one dealer receives the bulk of the catch, it is recommended that the working group and assessment team work closely with industry to collect better statistics of effort, standardised effort and size.

Each dealer over time, and within a year, uses different code names for the size categories, which is an important data source within the assessment. **Despite the uncertainties, these data have been combined in an appropriate way, especially since the year classes appropriately follow from one year to the next.**

A key feature of the age structure was that the catch consisted of a small set of age classes, with few to no extra large animals in the catch. The price was highest for the large category, but drops for the extra large. Although **the working group and assessment team appropriately addressed the uncertainties in the size data**, this remained an issue. The working group showed that the unclassified column consisted of mostly large-medium animals based on discussion with the industry and the average price per grade. Furthermore, some extra large animals were being put in the large category. **The final size and age data set used in the assessment and developed by the WG is supported.** Although some miscoding is likely, the group had undertaken as much as can be done with the present data.

There were several features of the size data that impact on the assessment. These included that the early Turner length-frequency data were available in coarse 5 cm units and that the 1983 data showed smaller animals appearing in the catch data, yet no data were available for the following years to confirm whether this was a recruitment pulse or a change in the fishery (e.g. through a selectivity or spatial coverage change).

Catch-at-age data also had to be pooled to fill in missing data in certain years. Although some interpolation methods may be possible, the present method is supported.

ToR 3 For the depth zone occupied by tilefish, examine the relationship between bottom temperature, tilefish distribution and thermal tolerance.

<u>This ToR was partially met</u>

There were very few data for tilefish, as they were not readily caught in the bottomfish surveys (e.g. 138 fish in 1968-2012 NEFSC spring survey). These data confirm their distribution on the shelf break, in temperatures from 4 to 15°C. The winter surveys for 1992-2007 showed a similar distribution pattern and temperature range, although the distribution is within slightly colder areas as well. Even fewer fish are caught in the fall surveys – 47 fish (1968-2012). A simple calculation of the probability of occurrence showed that the distribution implied by the data is in slightly colder waters than reported in the literature.

Since there were few survey data, only attempting to obtain insights about broad distribution and temperature profiles at a large temporal scale is supported.

However, from the literature, it is known that tilefish are potentially vulnerable to temperature intrusions of cold water into the area they inhabit. **The partially completed ToR finding is due to the GLM standardisation of catch rate not including large-scale environmental and climatic drivers, such as the North Atlantic Oscillation index.**

ToR 4 Use assessment models to estimate annual fishing mortality and stock size for the time series, and estimate their uncertainty. Include a historical retrospective to allow a comparison with previous assessment results.

<u>This ToR was met</u>

The assessments and analyses provided by the WG and assessment team were very comprehensive and rigorous. The analyses were completed by providing tests on model type, data and assumption sensitivity within the models and retrospective analyses. This work is of a high standard.

Three models are provided for review – surplus production model (ASPIC), an age and length forward projection model (SCALE) and an age model (ASAP). Much work by the group had been undertaken to develop age and size models, as the ASPIC model had previously been shown to have major issues fitting to these data.

For SARC 58, the ASPIC model revealed extreme retrospective patterns when compared to historical model formulations and runs. There was also a strong retrospective pattern for a 1-year peel. The model was unable to fit the strong cohorts shown within the fishery. The model was sensitive to whether the catch rate data were treated as a single or separate series.

The SARC 58 working group did not support the latest ASPIC runs for stock status determination, as the issues of process error still had not been resolved. As such, the ASPIC model was not supported. This decision is appropriate given the lack of model fit to the LPUE data and that it did not identify the strong cohorts entering the fishery.

The SCALE model did not rely on age-specific data on a yearly basis, which pointed to this model potentially been well suited for application to this fishery. The SCALE model estimated numbers at age and CPUE series in weight. The updated age data were used to estimate growth for the SCALE model – the result was growth similar to the growth curve estimated by Turner, especially at the lower ages.

The model estimated recruitment (R) in year 1 and variation in R relative to this initial R value. Since there is not an index of R, a flat line was assumed and the model estimated the deviation from this assumption. This method of estimating R was used in both SCALE and ASAP and is supported.

There was some uncertainty as to which M value should be inputted to the SCALE model. A value of 0.1 was previously used, but the Turner M values were 0.1, 0.15 and 0.2. The SARC 48 SCALE used M of 0.1 for the sexes combined. **The SARC 58 working group concluded that a range of M values from 0.1 to 0.15 is appropriate. This range is supported. Based on virgin length distributions from the SCALE model, an M of 0.15 is considered most appropriate.**

The SCALE model appropriately revealed R events that feed through the fishery. Retrospective patterns for the final SCALE run showed some patterns in F for the 2008 and earlier peels, but reasonably good retrospective test results for recruitment and total numbers. However, the SCALE model assumed flat topped selectivity and could not fit other formulations, especially dome-shaped selectivity. There was strong evidence from the catch-at-age structure and maximum age from growth studies that dome selectivity should be tested as an option. **The move away from the SCALE model to using the ASAP model as a basis for the base model is supported. The stated reasons of lack of data (mid 1990s, no recruitment index) in the model, issues with estimated selectivity by block, and dome or year class targeting were appropriate for this move.**

ASAP is a well-established age structured assessment model (Legault and Restrepo, 1998) that has been peer reviewed and undertaken extensive simulation testing. As a result, this review is about its application to tilefish. **For the information available and the assumptions that need to be tested, ASAP is the best model of the three made available to the SARC 58 panel.** An ASAP formulation was the final proposed base case for the SARC 58 panel to review. The key characteristic of ASAP is the ability to test for dome shaped and other selectivity patterns. Several sensitivity tests were undertaken by the working group and assessment team and some of these results were:

1) Plus group and age-length key
    a) catch-at-age to 20+ years using year specific expansions for years that age data are available,
    b) catch-at-age to 20+ using a pooled age-length key for all years,
    c) catch-at-age to 10+ years using year specific expansions for years that age data are available,
    d) catch-at-age to 10+ using a pooled age-length key for all years.

These sensitivity tests showed that there was little sensitivity to the choice of 10+ or 20+ years being assumed for the plus group age or to pooling the age length key (except in the early periods where there are data gaps). The 20+ dome shaped formulations had more convergence issues than the 10+ ones. For this reason, **the age-length key that used pooled data was used as the base case. This decision is supported.**

The F and SSB sensitivity to the age-length key was used as a basis for proposing overfished and overfishing proxies rather than model based values. This decision is also supported.

2) Natural mortality of 0.1 or 0.15.

As expected, there is some sensitivity to input M. The use of a value of 0.15 in the base model is supported given findings from SCALE and ASAP.

3) Weighout and VTR data
    a) Combining the Weighout and VTR data into a single series or keeping separate,
    b) Including 1991-1994 data in the VTR series.

The impact was similar to SCALE with an increase in retrospective pattern for the combined run. **The base model proposed to keep the series separate and use the 1991-1994 VTR data – this decision is supported.**

4) Using 2012 or 2013 as the terminal year (age data runs to 2012, catch data to 2013)

The model fit was not sensitive to the terminal year and yet forward projections required these last few years' age structure. Ending the catch series together with the age data is supported.

5) Selectivity
   a) Dome shaped versus flat topped,
   b) Different formulations of dome shaped for 10+ and 20+,
   c) Fixed or estimated components of the dome shaped selectivity,
   d) Estimated as a time invariant series or two sets of selectivities,
   e) If two separate series, the new period starting in 1982 or 1983.

Firstly, the dome shaped selectivity pattern to be estimated was of course dependent on the plus group year decision. Given the choice of 10+, three dome-shaped selectivity patterns were tested. This was a key assumption within the model and resulted in extensive discussion within the panel. To facilitate this discussion, population age structure for different runs was provided during the review.

The SARC 58 working group's argument of dome shaped selectivity is:

- **Tilefish is a long lived species of more than 30 years, yet the catch ages are of predominantly 4-6 years old animals caught by a small fleet of mostly 4 vessels in a small area within the distribution.**
- **Assuming flat-topped selectivity resulted in a very skewed estimated population structure showing almost no animals larger than 8 years old in the population and a high F. These numbers are not reflected in the raw numbers of fish aged.**
- **The population is spatially distributed with larger animals offshore of where the fishery occurs that was harder to fish and requires stronger gear.**
- **The price of XLarge animals is lower than the other sizes.**

For these reasons, the working group agreed that the very few animals estimated in the older age structure for flat-topped runs are unreasonable and a sign of flat-topped selectivity (through gear and spatial separation). However, further work on industry co-operative surveys is essential and recommended to further support this argument, as it is more conservative to select flat-topped selectivity.

Generally, the 10+ and 20+ formulations show similar results with respect to selectivity blocks. This issue arises due to the lack of length data from 1983-1994, and because the last year of the Turner data (1993) is unusual. This lack of subsequent data makes it difficult to determine whether smaller animals have entered the fishery or the selectivity of the fishery has changed, and therefore **the working group proposal to estimate two selectivity blocks starting in 1983 is supported.**

6) Fit to the 1974 length structure

Although it is unusual to remove data, this can be justified because missing data and reconstruction of data from different sources often means outliers are difficult to objectively assess. This data set is also a small sample size. **The removal of the unusual 1974 length structure is reasonably robust and supported.**

7) Length frequency effective sample size

The value used is supported.

The final proposed model run 27b is therefore accepted.

In the future, it is recommended that model tests of SCALE and ASAP be continued (for model structure tests), but that ASPIC is not an appropriate model for species with variable recruitment and should not continue to be applied to tilefish.

ToR 5 State the existing stock status definitions for “overfished” and “overfishing”. Then update or redefine biological reference points (BRPs; point estimates for BMSY, BTHRESHOLD, FMSY and MSY or for their proxies) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the “new” (i.e., updated, redefined, or alternative) BRPs.

This ToR was met

Prior to SARC 58, the existing stock status determination was based on an ASPIC model, which is now shown as inappropriate. **Given a) the lack of a complete data set especially the age-length key and model sensitivity to this key and b) sensitivity to other inputs such as the plus group, the use of proxy BRPs are supported.**

Spawning and Yield per recruit curves against fishing mortality show that using an F reference point of F40% (or more conservative) is appropriate. Since the 905mt quota had resulted in the stock having increased, recent fishing mortality rates have therefore allowed stock recovery. Since the implementation of the constant quota, F values have averaged 0.367, which corresponds to an F25% proxy value. The latter value was therefore proposed as the new overfishing reference point. **The F25% reference point is accepted.**

Long-term projections at this F proxy value (with uncertainty) were then used to calculate the SSBmsy proxy. **This method of calculating the resultant SSBmsy proxy is supported.**

Under the constant quota, the stock appeared to have produced strong recruitment classes and there was no indication of stock decline or collapse. Since the stock has been increasing over the past decade, the new BRP proxies are likely to be conservative.

ToR 6 Evaluate stock status with respect to the existing ASPIC model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.

a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
b. Then use the newly proposed model and evaluate stock status with respect to “new” BRPs and their estimates (from TOR-4).

This ToR was met

The new modelling approach is an improvement on previous models. It was comprehensively tested and uncertainty (model structure, assumption/input and estimation) was well accounted for.

These model runs cannot be compared with ASPIC output as they are very different, but ASPIC was shown to be a poor model choice for tilefish.

Based on ASAP 27b output and the new proxies, the working group position that overfishing is not occurring and the stock is rebuilt is supported.

ToR 7 Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

a. Provide numerical annual projections (2-3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

This ToR was met

Four versions of forward projections were provided before and during the panel review:

1. Using the Fmsy proxy of F25%,
2. Using a constant quota of 905 mt,
3. Using a constant quota of 800 mt, and
4. Adjusting the recent age structure to account for retrospective patterns in recruitment residuals due to the lack of recent early age data to estimate recent recruitment accurately.

Uncertainty was included in these projections.

It is appropriate that the estimated recruits in the recent years be empirically adjusted using information about average recruitment over the whole series and the retrospective patterns.

This method of adjusting recent recruitment and using the Fmsy proxy for forward projection was accepted by the panel.

The conclusions of low probabilities of overfishing and that the stock is not overfished is supported.

ToR 8 Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

This ToR was met

The WG reported on their progress with regard to previous recommendations. Much work has been undertaken in the intervening period including work on data uncertainty, obtaining ageing information of commercial tilefish catches and the use of additional models. These ageing data were essential and their collection should be continued. The SARC 58 WG also provided two new research recommendations:

- Develop an industry-based survey that would sample a greater proportion of the tilefish distribution, and
- Increase age-at-maturity sampling.

The industry-based sampling is supported and is seen as a high priority. This sampling should try to emphasise regions inshore and offshore of the present fishing grounds to include smaller and larger fish. However, the latter recommendation seems to have been addressed in a recent tilefish maturity study (McBride et al. 2013).

A few further recommendations were made:

- **The WG should investigate the potential effect of large scale environmental and climatic drivers on recruitment,**
- **Since 1 dealer records most of the catch and 4 fishers undertake most of the catch, work with this dealer and the industry for more accurate size data should be undertaken,**
- **If possible, undertake some form of age validation especially with regard to the important maximum age assumption.**

4.3 C. Northern Shrimp Terms of Reference

Background

Northern shrimp *(Pandalus borealis*) in the Gulf of Maine is at its southern most extent, but is thought to be a discrete stock. It prefers cold basins in depths of 90 to 180m on soft bottoms. The Atlantic States Marine Fisheries Commission (ASMFC), an intestate compact, manages the fishery. It has a complex biology characterised by a strong inshore (winter) and offshore (summer) migration pattern, and starting life as a male and changing to a female in about their third year of age. They live in waters ranging from 0-5°C. The spring temperature has been shown to be important for recruitment, with cold water better for recruitment. The 2012 sea surface temperature has been shown to be at its highest levels on record and similar to those temperatures in the 1950s when northern shrimp population size was low. It is an open access fishery with the fishery controlled by a complex set of input and output controls.

ToR 1 Present the Gulf of Maine northern shrimp landings, discards, effort, and fishery-independent data used in the assessment. Characterize the precision and accuracy of the data and justify inclusion or elimination of data sources.

This ToR was met

Commercial landings were provided from the 1960s to present from the NMFS database on dealer reports, harvester report data, dealer report data and Vessel Trip Reports (VTR). The 2013 data were considered preliminary. The earlier data were less complete than those collected more recently, but it was difficult to determine to what extent this disparity had occurred. Sensitivity runs were undertaken in the assessment to address this issue. Annual landings were characterised by being highly variable, ranging from a few hundreds of tonnes to more than 10,000 mt. Maine forms the largest part of the landings and the highest proportion of the catch occurs in February. There are two main gear types that are used to catch northern shrimp, being trap and trawl. Trawl gear catches over 80% of the landings. **Based on several sampling studies, the discards were correctly assumed to be zero.** Size, sex and stage information of landings was available, showing that the bulk of the catch consisted of ovigerous and spent females that were aged 4-5 years old.

Catch rate data would be influenced by the migration of northern shrimp, but also the complex input (such as daily catch days) and output (Total Catch and Total Allowable Catch) management system. Catch rates were calculated as either in units of mt/trip (since 1965), or lbs/hr (since the 1990s). Catch rate data were not used in the assessment since they were not deemed as a reliable indicator of biomass, especially when compared with the summer survey data. The lack of a standardised catch rate or effort series in the assessment is commented on in ToR 2 and 6 below.

Thus, the fishery independent survey data were a crucial data set. Five surveys were considered: the ASMFC summer shrimp survey, the NEFSC spring and fall bottom trawl survey, the Maine-New Hampshire Inshore trawl survey, State of Maine/DMR summer shrimp survey. Only two of these were used in the benchmark assessment, the ASMFC summer shrimp survey and the NEFSC Autumn Trawl survey (but the latter are two separate series due to a vessel change in 2009). The excluded surveys had limited ranges compared to actual northern shrimp distribution, had high variability or were influenced by confounding effects of inter-annual variation in the timing of post-hatch female northern shrimp. **This decision to presently only use these two survey sets is supported, but further investigation of the utility of the other surveys using model based approaches standardised for environmental effects should be investigated.**

The confidence intervals for the included surveys are reasonably narrow (except for the 2006 year, which was an unusual year with extremely large indices from both surveys). **These surveys are a good index of abundance. The survey size composition tracked stage and pseudo-age classes well and is useful for the assessment.**

ToR 2 Estimate population parameters (fishing mortality, biomass, and abundance) using assessment models. Evaluate model performance and stability through sensitivity analyses and retrospective analysis, including alternative natural mortality (M) scenarios. Include consideration of environmental effects where possible. Discuss the effects of data strengths and weaknesses on model results and performance.

<u>This ToR was not met</u>

Past assessments preferred the delay-difference model (called the Collie-Sissenwine Analysis, CSA, model), with an additional assessment being ASPIC. For the SARC 58 assessment, a new model was developed – a statistical catch-at-length model (UME). **This model is a great advance on those used before, with a length-based model being an appropriate step forward for hard-to-age species such as northern shrimp. Unlike the other models provided to SARC 58, the UME model has not been independently simulation tested and it is recommended that this process be undertaken in the near future.**

Furthermore, the CSA model was appropriately updated to include a formal likelihood framework that was able to incorporate multiple indices of abundance.

The work undertaken to advance the assessments from previous runs is greatly appreciated and supported.

Several methods of calculating natural mortality were attempted. **This work was rigorously undertaken, in that it not only used different methods of calculating M, but also looked at age-constant, age-varying, length-varying and time-varying methods. This work was comprehensive.**

In order to include time varying natural mortality as a test within the UME and CSA models, a predation-scaled time-varying index was developed from predator diet samples that contained Pandalids in their diet. This index was then based on the predator biomass indices and the frequency of Pandalids in the diet of each predator, and was used as an input to scale M in each year of the assessment. **This Prawn Predation Index (PPI) (although based on small samples) appeared to be a good index and further research on refining this work is supported (see comments regarding its use in the models below).**

The UME model was set up using an annual time step, starting from 1984 to present. A seasonal model was attempted, but this did not converge and was difficult to parameterise. The temporal extent of the model was based on the time series of the shrimp surveys. The model was set up with:

- two sexes (females and non-females),
- three fleets with logistic selectivity (mixed 1984-1999, trap 2000-2013, trawl 2000-13),
- two growth stanzas – cold <2000 and warmer ≥2000 – where the von Bertalanffy growth parameters Kappa and L_∞ were held constant between the two periods but the variance around these parameters, estimated within the model, were assumed to be different for the two periods,

- U-shaped M for the base model, with constant M and PPI tests, and
- Equally weighted likelihood components.

The use of the U-shaped M as a function of size in the proposed base model needs much greater justification through observation or other evidence than that provided and its use as the base model is not supported. It was calculated using three different methods with arbitrary length transition points:

- Small northern shrimp M was calculated using Lorenzen (1996) (scaled to reflect a 6-year longevity animal),
- The middle sized northern shrimp M used the Rinaldo (1976) M estimate (0.2: lower than the constant M values), and
- A higher constant M of 0.75 for larger northern shrimp (within the constant M values).

The latter value was argued based on there being a terminal moult. Although there may be anecdotal evidence for a terminal moult, there is no direct evidence for this. The transition points between the different M values were likely to create unusual characteristics in the model, especially in combination with the growth matrix and the sex change matrix. **Without much clearer evidence and tests (see ToR 6), the simpler constant M is recommended.**

Although the text stated that the model fit the data well, this conclusion was not supported. There was a clear model misfit for the key summer survey index and the size distributions. Residual patterns were strong and the terminal year residual was large with the terminal recruitment overestimated. This is partly due to the equal weights within the likelihood driving the model to fit the data with the lowest CVs, in this case the catch data. However, despite the fact that the catch data are much more accurate, from a total likelihood point of view, this data set was given much more overall weight in the likelihood than the much more important independent index of abundance. **Sensitivity tests to different weightings were undertaken, including changing the survey weightings. This is a practice that is often not standard and adds enormously to the rigour with which this assessment was undertaken. Furthermore, these sensitivity tests also provided a key management parameter being terminal SSB/Bmsy.** These results show that the model results are very sensitive to values of M (expected) and to survey likelihood weight. **Retrospective tests also showed that the model had strong retrospective patterns in key management statistics.**

The panel requested several UME sensitivity tests (with a new base model of a constant M of 0.5) during the review:

1. Likelihood weight test 1: double surveys base model weight, halve catch base weight, unchanged survey size composition base weight
    a. New residual plots (aggregated size composition residuals) of the base case show these fit the size composition data better than adjusted likelihood test 1:
    b. Changing the weighting had not removed the issue of fitting to the 2006 high recruitment value;
    c. The fits had also not been improved regarding the survey data.
    d. **The constant M had lower parameter space and did not affect the model fit or results. Although, based on natural history, there may be a case for U shaped M, from the perspective of parsimony the U-shaped M is not supported.**
2. Likelihood weight test 2: same as test 1 but the CVs for the surveys are those calculated (i.e. clearly defining the role of the CV component in the likelihood), rather than as a product of iterative weighting from a previous model output.
    a. This again had no dramatic effect of the residual pattern
3. Treat the 2006 data point as missing
    a. Still got a strong residual pattern and a large 2013 residual

Thus, in conclusion, the UME size model was an appropriate model type to use on hard-to-age species. However, the model was unable to fit the data appropriately, had strong retrospective patterns in the spawner biomass estimate and derived little additional value (compared to the

CSA) from including the size composition data. For this reason, the use of the UME model for northern shrimp at this stage is not supported. Investigations of adding correlates to further explain data are likely to be needed.

Two alternative models were also provided – CSA and ASPIC. The model structure of ASPIC was unable to fit the high variability in the survey indices even though it was able to undertake runs with a longer time series. The ASPIC model was particularly affected by the post 2005 data and showed very strong retrospective patterns. **Presently, the ASPIC formulation is inappropriate for the large variability in the post 2005 data. It is therefore not recommended for present use on northern shrimp.**

The Collie-Sissenwine Analysis (CSA) model has been peer-reviewed and accepted for use on northern shrimp by SARC 45. This two-stage model has value in that it is able to capture large variability in recruitment (something the ASPIC model was usually unable to achieve), but is unable to use size data. **CSA is therefore often a good compromise in terms of being an intermediate parameter model to ASPIC and UME. The use of this model type for northern shrimp is supported as it presently provides a best compromise to UME (until the peer review issues and lack of the size data providing additional information are resolved).**

Despite the above, CSA was still unable to appropriately fit the index data given the poor residual fits. Adding time varying M in the form of the PPI removed the strong retrospective patterns. This showed that, presently, the CSA run using the PPI performed better than the UME (at least with respect to retrospective patterns).

The panel requested several CSA sensitivity tests during the review:

1. M=0.5 or M=PPI; likelihood weightings lambda =(double ASMFC survey indices weight; unchanged for NEFSC survey index weight; half for catch weight)
    a. **This further supports that, based on Mohn's rho retrospective tests, the M= PPI was best**
    b. The adjusted lambdas run compared to the base run was much better from Mohn's rho point of view as well as likelihood values
2. The above adjusted lambdas, catch CV=0.2 (instead of 0.05), M=PPI
    a. This model performed only slightly better with respect to Mohn's rho
3. M=PPI, C=0.2 catch, Catch lambda from 0.5 to 0.05
    a. The best survey index residual pattern test was with a catch lambda of 0.01. The positive residuals decreased in the earlier part of the time series, but remained large over the 2006 period. The model fit to the catch increased the estimated catch in 2006 compared to observed catch, further demonstrating the incongruity of the catch and survey series data.
4. Increasing combinations of survey lambdas and decreasing catch lambdas
    a. All highlighted the poor residual fit around 2006, which cannot be removed.
    b. **Although component values of the likelihood changed with the different weightings (as expected), the total likelihood value remained the same even at the extreme tests of weighting for the catch or for the survey indices. This means one would not be able to use objective model fit methods to choose the most appropriate likelihood weighting.**
5. As a simple test, plot the ratio of Cy+1/Cy and SSBy+1/SSBy for the extreme tests (these are related to key management measures)
    a. These plot results were very different for the two extreme likelihood weight tests, meaning that the management advice is very sensitive to the likelihood component weightings.

In conclusion, the CSA model applied to northern shrimp was unable to fit the data appropriately and the model output was very sensitive to likelihood weightings. It is therefore not supported for determining fishing mortality and stock status.

As an overall conclusion:

1. **A major issue with respect to any of the model fits was the sudden increase in recruitment in 2005 and 2006 shown in both surveys, followed by a sharp decrease in the indices thereafter (without large catches). Similarly, recent indices were extremely low with reasonably low catches. This means that the model fit issues described in this review are not due to the assessment team performance or the actual model program, but rather the data provided to the model being hard to explain without some other correlate information (e.g. temperature). Both the CSA and UME models are flexible enough to capture large recruitment variability.**
2. **There does not seem to be any benefit to adding length data, as the UME model did not fit the size data well.**
3. **There was also no sensitivity of the UME model to a U-shaped M or flat M, so a flat M should be used.**
4. **The ASPIC model did not fit the high variability in the newly added data and would remain unable to do so because it does not have enough parameter space needed for this flexibility.**
5. **Including annual changes to M through using the Predation index, PPI, benefited the CSA model and allowed greater flexibility and better model fits (especially in terms of decreasing retrospective patterns).**
6. **The CSA at this stage is a good compromise between model parsimony and flexibility.**
7. **However, fishing mortality and stock size estimates were not robust to the relative weights of the different indices.**
8. **Therefore, none of the model formulations applied to northern shrimp are robust enough to be used for management advice.**

ToR 3 Update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, SSBMSY, FMSY, MSY). Evaluate stock status based on BRPs.

This ToR was not met. The work was undertaken, but the output cannot be used for management advice

Given that the panel did not accept any of the model formulations as applied to northern shrimp, stock status and the proposed SARC 58 BRPs based on these model outputs cannot be determined. However, this fishery has valuable survey indices with low CVs that appeared to be proportional to stock size. **Stock status should therefore be determined using a combination of catches and survey indices.** In that regard, the recent survey indices indicated that the stock abundance is very low.

ToR 4 Characterize uncertainty of model estimates of fishing mortality, biomass and recruitment, and biological reference points.

This ToR was not met. The work was undertaken, but the output cannot be used for management advice

Several forms of uncertainty were addressed by the SARC 58 WG. They articulated the model uncertainty well. **However, given that the panel did not accept the proposed assessment models, appropriate evaluation of uncertainty in the model estimates and BRPs was not possible.**

ToR 5 Review the methods used to calculate the annual target catch and characterize uncertainty of target catch estimates.

This ToR was not met. The work was undertaken, but the output cannot be used for management advice

Given the panel did not accept the proposed assessment models, appropriate calculation of the annual target catch and associated uncertainty was not possible.

ToR 6 Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made before the next benchmark assessment.

This ToR was met

A comprehensive list of research priorities was provided. These were divided into short and long-term priorities for different categories of work (e.g. fishery-dependent priorities) and, within these, the relative importance of these was provided. This provided a clearly prioritised list of research recommendations. **These research priorities are supported.**

Given the importance of the shrimp specific surveys, these should be continued.

However, it is recommended that a few priorities should be updated or and new research projects should be added to the list. The most obvious addition is undertaking work that would provide a robust assessment model.

- **This work should concentrate on the CSA model, in the first stage.**
- **To progress resolution of the difficulties with the data, it is essential to investigate the possibility of adding correlates (such as temperature) to the model. This should be a high priority.**
- The correlations between the CPUE inshore, CPUE offshore and CPUE total show that these were strongly correlated (>0.8), whereas that between CPUE total and landings is poor (0.2). This highlights that the catch data were confounded by management measures and a low catch was not always a reflection of low catch rate, and vice versa. The correlations between CPUE total, and the shrimp survey index of animals larger than 22mm CL, the shrimp survey total index and the NEFSC index were above 0.9 pointing to potential usefulness of these data. However, this latter statement needs to be moderated by the fact the catch rate data do not easily correspond to all the high and low values in the survey indices (see Figure 1). **Although there are issues with effort and catch rate, a standardised series is likely to greatly improve the contradiction between the catch and survey indices – even if the catch rate data are added with low weight.**

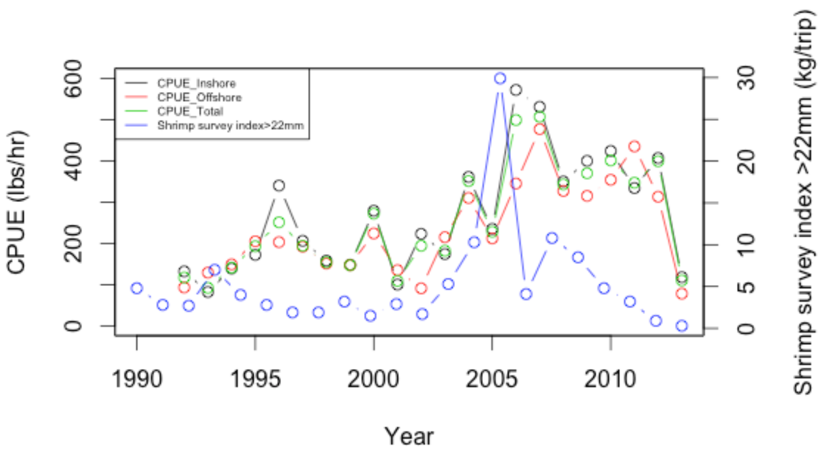

Figure 1: Northern shrimp CPUE of animals and survey index greater than 22 mm CL over time.

- **Given the relevance of using a size-based model for hard-to-age animals, it is recommended that simulation testing of the UME model be undertaken.**
- **If support for the U-shaped M is needed, some form of independent verification of a terminal moult is needed through tagging or some other study.**
- **Obtaining appropriate BRPs in a data limited environment (through simulations such as management strategy evaluation) should be investigated.**

ToR 7 Based on the biology of species, and potential scientific advances, comment on the appropriate timing of the next benchmark assessment and intermediate updates.

<u>This ToR was met</u>

Given the issues highlighted by the panel and this report regarding the assessment, the Panel recommended a staged approach of:

- Initially basing management advice using observed patterns of the survey indices and catches,
- Once an updated assessment (whether CSA or UME) is completed, a benchmark assessment should be undertaken.

5 Conclusions and Recommendations

The butterfish, tilefish and shrimp assessment teams and associated working groups have achieved considerable advances in characterising the underlying data required in the assessments and describing uncertainty. Specifically, these groups took great care in fully specifying and testing for different kinds of assessment uncertainty: model structure uncertainty through the use of different models with very distinct assumptions, sensitivity of each model to input and parameter uncertainty, sensitivity of the models to new data through retrospective tests. This should be seen as a model for use by other groups.

For *butterfish*, all Terms of References (ToRs) were fully met and the assessment could be used for management advice. Several advances have been made since the last assessment that greatly enhanced the SARC 58 reviewed body of work. Research that linked oceanographic information with the survey catchability should be continued into the future. Small changes to the proposed base model were tested, and this updated version was accepted by the Panel. The accepted base model (without the habitat index and the spring survey) yielded low estimated F rates, and a relatively healthy stock abundance. There was consensus among the Panel members that this slightly modified base model can be used for management purposes. Overfishing is not occurring and the stock is not overfished.

For *golden tilefish*, all ToRs were met. Since the last supported assessment, there have been several advances in the input data and applying different new assessment models to tilefish. The final base model was accepted by the Panel, but with a minor change in the method of forward projecting the age structure. The final accepted model results showed that overfishing is not occurring and that the stock has been rebuilt.

Despite the rigorous effort of the assessment team, not all the *northern shrimp* ToRs were met. The assessment models were not robust to tests undertaken during the review and should not be used for management advice. Since the last assessment, large advances were made especially the delivery of a new additional assessment model (UME – a size-structured assessment designed for hard-to-age animals such as shrimps). The different models were unable to appropriately characterise the data and provide robust estimates of variables important for management advice. This result with regard to northern shrimp is not a reflection on the assessment team, but rather a feature of the new data showing a large recruitment pulse that is not consistently followed through into the subsequent catch and survey data. Future work should concentrate on adding potential covariates to the model e.g. bottom temperature.

Research recommendations are provided against the relevant ToRs in the text.

With regard to the SARC 58 process, the meeting was well run and the review teams were open to discussion and willing to undertake further work. Their enthusiasm and professionalism was greatly appreciated.

6 References

Freeman BL, Turner SC. 1977. Biological and fisheries data on tilefish, *Lopholatilus chamaeleonticeps* Goode and Bean. NEFSC Tech Rep. 5.

Johnson FH, Lewin I. 1946. The growth rate of E. coli in relation to temperature, quinine and coenzyme. Journal of Cellular and Comparative Physiology 28:47-75.

Legault CM, Restrepo VR. 1999. A flexible forward age-structured assessment program. Col Vol Sci Pap ICCAT 49:246-253.

McBride RS, Vidal TE, Cadrin SX. 2013. Changes in size and age at maturity of northern stock of tilefish (Lopholatilus chamaeleonticeps) after a period of overfishing. Fish Bull. 111:161-174

Turner, SC. 1986. Population dynamics of and, impact of fishing on tilefish, *Lopholatilus chamaeleonticeps*, in the Middle Atlantic-Southern New England region during the 1970's and early 1980's. New Brunswick, N.J.: Rutgers University. Ph.D. dissertation.

Appendix 1 Bibliography of materials provided for review

6.1 Working Papers

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Report of Butterfish. Working Paper #1. SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Report of Tilefish. Working Paper #1.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Report of Northern shrimp. Working Paper #1.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Summary Report of Butterfish. Working Paper #2. SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Summary Report of Tilefish. Working Paper #2.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Summary Report of N. shrimp. Working Paper #2.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

6.2 Background papers

Cadrin SX, Boutillier JA, & Idoine JS. 2004. A hierarchical approach to determining reference points for Pandalid shrimp. Canadian Journal of Fisheries and Aquatic Sciences 61(8): 1373-1391. 19 p.

Chen Y, Hunter M, Vadas R, Beal B. 2003. Developing a growth-transition matrix for the stock assessment of the green sea urchin (*Strongylocentrotus droebachiensis*) off Maine. Fishery Bulletin 101:737-744. 8 p.

Chen Y, Kanaiwa M, & Wilson C. 2005. Developing and evaluating a size-structured stock assessment model for the American lobster, *Homarus americanus*, fishery. New Zealand Journal of Marine and Freshwater Research (Special issue on Lobster Biology and Management). 39:645-660. 16 p.

Clark SH, Cadrin SX, Schick DF, Diodati PJ, Armstrong MP, and McCarron D. 2000. The Gulf of Maine Northern Shrimp (*Pandalus borealis*) Fishery: a Review of the Record. J. Northw. Atl. Fish. Sci. 27: 193-226. 34 p.

Cotter J. 2009. Northeast Stock Assessment Review Committee (SARC 49) for Atlantic Surfclams in the US EEZ and Butterfish. 30 Nov – 3 Dec 2009. Woods Hole, MA. 49 p.

Link JS & Idoine JS. 2009. Estimates of Predator Consumption of the Northern Shrimp Pandalus borealis with Implications for Estimates of Population Biomass in the Gulf of Maine. North American Journal of Fisheries Management, 29(6): 1567-1583. 18 p.

McBride et al. 2013. Changes in size and age at maturity of the northern stock of Tilefish *(Lopholatilus chamaeleonticeps*) after a period of overfishing. Fish Bull. 111: 161-174. 14 p.

Marsh et al. 1999. The 1882 tilefish kill - a cold event in shelf waters off the north-eastern United States? Fish. Oceanogr. 8:1, 39-49, 1999. 11 p.

Northeast Fisheries Science Center. 2007. 45th Northeast Regional Stock Assessment Workshop (45th SAW): 45th SAW assessment report. US Dep Commer, Northeast Fish Sci Cent Ref Doc 07-16; 370 p.

Northeast Fisheries Science Center. 2009. 48th Northeast Regional Stock Assessment Workshop (48th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-15; 834 p.

Northeast Fisheries Science Center. 2009. 48th Northeast Regional Stock Assessment Workshop (48th SAW) assessment summary report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-10; 50 p.

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-01; 41 p.

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-03; 383 p.

Northeast Fisheries Science Center. 2009. Summary Report of the 49th Northeast Regional Stock Assessment Review Committee (SARC 49). 30 Nov – 3 Dec 2009. Woods Hole, MA. 37 p.

Northeast Fisheries Science Center. 2013. Butterfish model diagnostics for the ASAP3 base model which served as the basis for developing the final model for SAW58 in 2014. 73p.

Richards RA, Fogarty MJ, Mountain DG, Taylor MH. 2012. Climate change and northern shrimp recruitment variability in the Gulf of Maine. Marine Ecology Progress Series 464:167-178. 12 p.

Smith MT. Northeast Regional Stock Assessment Review Committee (SARC 49). 2009. Independent reviewer's report on the 2009 Atlantic Surfclam and Butterfish Benchmark Stock Assessments. December 2009. 30 Nov – 3 Dec 2009. Woods Hole, MA. 51 p.

Sparholt H. Northeast Regional Stock Assessment Review Committee (SARC 49). 2009. SARC 49: Butterfish and Atlantic surfclam Benchmark Stock Assessments. 30 Nov – 3 Dec 2009. Woods Hole, MA. 33 p.

Appendix 2 Copy of the Statement of Work

Statement of Work

58th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for butterfish, tilefish, and northern shrimp

Statement of Work (SOW) for CIE Panelists

(including a description of SARC Chairman's duties)

BACKGROUND

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are independently selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description: The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development (SAW Working Groups or ASMFC technical committees), assessment peer review, public presentations, and document publication. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fishery management in the northeast region.

Brief description of the science to be peer reviewed, and its relevant importance:

The purpose of this meeting will be to provide an external peer review of benchmark stock assessments for **butterfish**, **tilefish**, and **northern shrimp**. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results form the scientific basis for fishery management in the northeast region.

OBJECTIVES

The SARC review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the SSC of the New England or Mid-Atlantic Fishery Management Council. The SARC panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Duties of reviewers are explained below in the "**Requirements for CIE Reviewers**", in the "**Charge to the SARC Panel**" and in the "**Statement of Tasks**". The stock assessment Terms of Reference (ToRs) are attached in **Annex 2**. The draft agenda of the panel review meeting is attached in **Annex 3**. The SARC Summary Report format is described in **Annex 4.**

Requirements for the reviewers: Three reviewers shall conduct an impartial and independent peer review of the striped bass and summer flounder stock assessments, and this review should be in accordance with this SoW and stock assessment ToRs herein. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include statistical catch-at-age, state-space and index methods. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points that includes an appreciation for the varying quality and quantity of data available to support estimation of Biological Reference Points. SARC 58 will address fishery stock assessments of **butterfish, tilefish,** and **northern shrimp.** For shrimp and butterfish, experience in the following is desirable: assessment of short-lived species, stocks where the environment and environmental change can impact recruitment and availability in research surveys. Specifically for tilefish: experience with assessments based on commercial catch per unit of effort.

PERIOD OF PERFORMANCE

The contractor shall complete the tasks and deliverables as specified in the schedule of milestones within this statement of work. Each reviewer's duties shall not exceed a maximum of 16 days to complete all work tasks of the peer review described herein.

Not covered by the CIE, the SARC chair's duties should not exceed a maximum of 16 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation).

PLACE OF PERFORMANCE AND TRAVEL

Each reviewer shall conduct an independent peer review during the panel review meeting scheduled in Woods Hole, Massachusetts during dates of January 27-31, 2014.

STATEMENT OF TASKS

Charge to SARC panel: During the SARC meeting, the panel is to determine and write down whether each stock assessment Term of Reference (ToR) of the SAW (see **Annex 2**) was or was not completed successfully. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. **If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted.** Where possible, the SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment Term of Reference of the SAW.

If the panel rejects any of the current BRP or BRP proxies (for B_{MSY} and F_{MSY} and MSY), the panel should explain why those particular BRPs or proxies are not suitable, and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs or BRP proxies are the best available at this time.

Each reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, phone number, FAX number, and a CV suitable for the public) to the COR, who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Foreign National Security Clearance: The reviewers shall participate during a panel review meeting at a government facility, and the NMFS Project Contact will be responsible for obtaining the Foreign National Security Clearance approval for the reviewers who are non-US citizens. For this reason, the reviewers shall provide by FAX (or by email if necessary) the requested information (e.g., 1.name [first middle and last], 2.contact information [address, telephone number], 3.gender, 4.country of birth, 5.country of citizenship, 6.country of permanent residence, 7.whether there is dual citizenship, 8.country of current residence, 9.birth date [mo, day, year], 10.passport number, 11.country of passport) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/.

Pre-review Background Documents and Working Papers: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the SARC chair and CIE reviewers the necessary background information and reports (i.e., working papers prepared by the SAW Working Group) for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements).

The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

(SARC chair)

Act as chairperson, where duties include control of the meeting, coordination of presentations and discussions, making sure all stock assessment Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For each assessment, review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)

For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. From a reviewer's point of view, determine whether each stock assessment Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point or BRP proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist. Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

Tasks after the panel review meeting:

SARC CIE reviewers:

Each CIE reviewer shall prepare an Independent CIE Report (see **Annex 1**). This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified above in the "Charge to SARC panel" statement.

If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report produced by each reviewer.

The Independent CIE Report can also be used to provide greater detail than the SARC Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

SARC chair:

The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was adequate to complete the stock assessment Terms of Reference of the SAW. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report (see **Annex 4**).

SARC chair and CIE reviewers:

The SARC Chair, with the assistance from the CIE reviewers, will prepare the SARC Summary Report. Each CIE reviewer and the chair will discuss whether they hold similar views on each stock assessment Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this SARC Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion.

The SARC Summary Report (please see **Annex 4** for information on contents) should address whether each stock assessment Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

DELIVERY

Each reviewer shall complete an independent peer review report in accordance with the SoW. Each reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each reviewer shall complete the independent peer review addressing each stock assessment ToR listed in **Annex 2**.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the Woods Hole, Massachusetts scheduled during January 27-31, 2014.
3) Conduct an independent peer review in accordance with this SoW and the assessment ToRs (listed in **Annex 2**).
4) No later than February 14, 2014, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and to Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each assessment ToR in **Annex 2**.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

December 16, 2013	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
January 13, 2014	NMFS Project Contact will attempt to provide reviewers the pre-review documents
January 27-31, 2014	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
January 31, 2014	SARC Chair and CIE reviewers work at drafting reports during meeting at Woods Hole, MA, USA
February 14, 2014	Reviewers submit draft independent peer review reports to the contractor's technical team for independent review
February 17, 2014	Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair *
February 21, 2014	SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)
February 28, 2014	Contractor submits independent peer review reports to the COR who reviews for compliance with the contract requirements
March 7, 2014	The COR distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on substitutions. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: The deliverables shall be the final peer review report from each reviewer that satisfies the requirements and terms of reference of this SoW. The contract shall be successfully completed upon the acceptance of the contract deliverables by the COR based on three performance standards:

1. each report shall be completed with the format and content in accordance with **Annex 1**,
2. each report shall address each stock assessment ToR listed in **Annex 2**,
3. each report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Upon the acceptance of each independent peer review report by the COR, the reports will be distributed to the NMFS Project Contact and pertinent NMFS science director, at which time the reports will be made publicly available through the government's website.

The contractor shall send the final reports in PDF format to the COR, designated to be William Michaels, via email William.Michaels@noaa.gov

Support Personnel:

William Michaels, Program Manager, COR

NMFS Office of Science and Technology

1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910

William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator

Northern Taiga Ventures, Inc.

10600 SW 131st Court, Miami, FL 33186

shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President

Northern Taiga Ventures, Inc. (NTVI)

22375 Broderick Drive, Suite 215, Sterling, VA 20166

RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Dr. James Weinberg, NEFSC SAW Chairman, NMFS Project Contact

Northeast Fisheries Science Center

166 Water Street, Woods Hole, MA 02543

James.Weinberg@noaa.gov (Phone: 508-495-2352) (FAX: 508-495-2230)

Dr. William Karp, NEFSC Science Director

Northeast Fisheries Science Center

166 Water St., Woods Hole, MA 02543

william.karp@noaa.gov Phone: 508-495-2233

Annex 1: Format and Contents of Independent Peer Review Report

1. The independent peer review report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).

2. The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs. For each assessment reviewed, the report should address whether each ToR of the SAW was completed successfully. For each ToR, the Independent Review Report should state why that ToR was or was not completed successfully. To make this determination, the SARC chair and reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the SARC Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not others read the SARC Summary Report. The independent report shall be an independent peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of this Statement of Work

   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: 58th SAW/SARC Stock Assessment Terms of Reference

(file vers.: 8/2/2013)

A. Butterfish

1. Characterize the commercial catch including landings, effort and discards by gear type. Describe the magnitude of uncertainty in these sources of data.

2. Characterize the survey data that are being used in the assessment. Describe the magnitude of uncertainty in these sources of data.

3. Characterize oceanographic and habitat data as it pertains to butterfish distribution and availability. If possible, integrate the results into the stock assessment (TOR-5).
4. Evaluate consumptive removals of butterfish by its predators. If possible, integrate results into the stock assessment (TOR-5).
5. Use assessment models to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a comparison with previous assessment results and previous projections.

6. State the existing stock status definitions for “overfished” and “overfishing”. Given that the stock status is currently unknown, update or redefine biological reference points (BRPs; point estimates for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY, or their proxies) and provide estimates of their uncertainty. Consider effects of environmental factors on stability of reference points and implications for stock status.
7. Evaluate stock status with respect to a newly proposed model and with respect to “new” BRPs and their estimates (from TOR-6). Evaluate whether the stock is rebuilt.
8. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
    a. Provide numerical annual projections (2 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment). Comment on which projections seem most realistic.
    b. Describe this stock’s vulnerability (see “Appendix to the SAW TORs”) to becoming overfished, and how this could affect the choice of ABC.
9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

B. Tilefish

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the magnitude of uncertainty in these sources of data.

2. Characterize commercial LPUE as a measure of relative abundance. Consider the utility of recreational data for this purpose. Characterize the uncertainty and any bias in these sources of data.

3. For the depth zone occupied by tilefish, examine the relationship between bottom temperature, tilefish distribution and thermal tolerance.

4. Use assessment models to estimate annual fishing mortality and stock size for the time series, and estimate their uncertainty. Include a historical retrospective to allow a comparison with previous assessment results.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY or for their proxies) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.
6. Evaluate stock status with respect to the existing ASPIC model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
    a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
    b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-4).
7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
    a. Provide numerical annual projections (2-3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
    b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
    c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.
8. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

C. Northern shrimp

1. Present the Gulf of Maine northern shrimp landings, discards, effort, and fishery-independent data used in the assessment. Characterize the precision and accuracy of the data and justify inclusion or elimination of data sources.

1.

2. Estimate population parameters (fishing mortality, biomass, and abundance) using assessment models. Evaluate model performance and stability through sensitivity analyses and retrospective analysis, including alternative natural mortality (M) scenarios. Include consideration of environmental effects where possible. Discuss the effects of data strengths and weaknesses on model results and performance.

2.

3. Update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, SSB_{MSY}, F_{MSY}, MSY). Evaluate stock status based on BRPs.

3.

4. Characterize uncertainty of model estimates of fishing mortality, biomass and recruitment, and biological reference points.

4.

5. Review the methods used to calculate the annual target catch and characterize uncertainty of target catch estimates.

5.

6. Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made before the next benchmark assessment.

6.

7. Based on the biology of species, and potential scientific advances, comment on the appropriate timing of the next benchmark assessment and intermediate updates.

7.

Annex 2 (cont.):

Appendix to the SAW Assessment TORs:

Clarification of Terms
used in the SAW/SARC Terms of Reference

Appendix to the Assessment TORs:

Explanation of "Acceptable Biological Catch" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 74, no. 11, 1/16/2009):

> *Acceptable biological catch (ABC)* is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of [overfishing limit] OFL and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*
>
> *ABC for overfished stocks.* For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect the annual catch that is consistent with the schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*
>
> NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)
>
> ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, [optimal yield] OY does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

Explanation of "Vulnerability" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 74, no. 11, 1/16/2009):

> *"Vulnerability.* A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce MSY and to recover if the population is depleted, and susceptibility is the potential for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Rules of Engagement among members of a SAW Assessment Working Group:

> Anyone participating in SAW assessment working group meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

Annex 3: Draft Agenda

58th Northeast Regional Stock Assessment Workshop (SAW 58)
Stock Assessment Review Committee (SARC) Meeting

January 27-31, 2014

Stephen H. Clark Conference Room – Northeast Fisheries Science Center

Woods Hole, Massachusetts

DRAFT AGENDA* **(version: 25 October 2014)**

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Monday, Jan. 27

10 – 10:30 AM

Welcome — **James Weinberg**, SAW Chair

Introduction — **Robert Latour**, SARC Chair

Agenda

Conduct of Meeting

10:30 – 12:30 Assessment Presentation (A. Butterfish)

TBD TBD **TBD**

12:30 – 1:30 PM Lunch

1:30 – 3:00 Assesssment Presentation (A. Butterfish)

TBD TBD **TBD**

3:00 – 3:15 Break

3:15 – 5:15 **S**ARC Discussion w/ Presenters (A. Butterfish)

Robert Latour, SARC Chair **TBD**

5:15 – 5:45 Public Comments (A. Butterfish)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Tuesday, Jan. 28

8:45 – 11 AM Assessment Presentation (B. Tilefish)

TBD TBD **TBD**

11:00 - 11:15 Break

11:15 – 12:15 SARC Discussion w/presenters (B. Tilefish)

Robert Latour, SARC Chair **TBD**

12:15 – 12 :30 Public Comments (B. Tilefish)

12:30 – 1:45 PM Lunch

1:45 – 4:15 Assessment Presentation (C. Northern shrimp)

TBD **TBD**

4:15 - 4:30 Break

4:30 – 5:45 SARC Discussion w/presenters (C. Northern shrimp)

Robert Latour, SARC Chair **TBD**

5:45 – 6:00 Public Comments (C. Northern shrimp)

7:00 **(**Social Gathering **)**

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Wed. Jan. 29

9:00 – 11:15 AM Revisit with presenters (A. Butterfish)

Robert Latour, SARC Chair **TBD**

11:15 – 11:30 Break

11:30 – 12:30 Revisit with presenters (B. Tilefish)

Robert Latour, SARC Chair **TBD**

12:30 – 1:30 PM Lunch

1:30 -2:30 (cont) Revisit with presenters (B. Tilefish)

Robert Latour, SARC Chair **TBD**

2:30 – 2:45 Break

2:45 – 5:15 Revisit with presenters (C. Northern shrimp)

Robert Latour, SARC Chair **TBD**

Thur. Jan. 30

8:30 – 11:30 Review/edit Assessment Summary Report (A. Butterfish)

Robert Latour, SARC Chair **TBD**

11:30 – 12:30 PM Lunch

12:30 – 2:45 Review/edit Assessment Summary Report (B. Tilefish)

Robert Latour, SARC Chair **TBD**

2:45 – 3:00 Break

3:00 - 5:30 Review/edit Assessment Summary Report (C. Northern shrimp)

Robert Latour, SARC Chair **TBD**

Friday, Jan. 31

9:00 AM – 5:00 PM SARC Report writing. (closed meeting)

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public, except where noted.

The NMFS Project contact will provide the final agenda before the meeting.

Reviewers must attend the entire meeting.

Annex 4: Contents of SARC Summary Report

1. The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the SARC chair and CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If the CIE reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

   The report may include recommendations on how to improve future assessments.

2. If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

8. The report shall also include as a separate appendix the assessment Terms of Reference used for the SAW, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 3 Panel membership or other pertinent information from the panel review meeting

Facilitator	Dr James Weinberg, NOAA (The United States of America)
Chair	Prof Rob Latour, Virginia Institute of Marine Science (The United States of America)
Members	Dr Catherine Dichmont, CSIRO (Australia)
	A/Prof Stewart Frusher, The University of Tasmania (Australia)
	Dr Ian Jonsen, Dalhousie University (Canada)

CONTACT US

t 1300 363 400
+61 3 9545 2176
e enquiries@csiro.au
w www.csiro.au

YOUR CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

Dr Cathy Dichmont
t +61 7 3833 5925
e cathy.dichmont@csiro.au
w www.cmar.csiro.au/